Bandolier 99

What do we think?
What do we know?
What can we prove?

Evidence-based health care **£3.00** **May 2002 Volume 9 Issue 5**

Subscribing to Bandolier

On page 8 of May's ***Bandolier*** you will find a subscription form for those of you who wish to continue to receive a monthly copy when NHS funding ceases after July. By filling in the form and returning it to us, your ***Bandolier*** will continue. Health service organisations and charities who can take ***Bandolier*** in bulk and distribute it themselves, or those who would like a bulk order for us to distribute, can contact us for preferential rates. The aim will always be to keep costs to the public sector to a minimum.

No ***Bandolier*** after July unless you take out a subscription.

Defining an agenda

Frequently-asked question number one when ***Bandolier*** takes to the road is how we choose topics for these pages, or what drives the agenda. The agenda is mostly driven by things that crop up, or begin to make sense, or where several bits of evidence come together by serendipity.

Take rheumatoid arthritis, for example. A really difficult topic which has always scrambled our brains, because every talk we sit through seems to add another layer of complex immunology, but not much to therapy or diagnosis. Now we have humanised monoclonal antibodies that can zap the inflammatory process, and mature NICE evidence with clinical outcomes we can just about get our heads around. Added to which a group of clever international rheumatologists provide a readable and sensible discourse on diagnosis and the benefit of early referral. Plus it's all done using high quality evidence methods. It's a gift.

Depression is another difficult topic. A good review of diagnostic methods with diagnostic features and classification lifts the gloom, even if a little. There is something that can be understood by anyone. Together with rheumatoid arthritis, we have two topics to add to a quality service structure that are useful and useable.

Old chestnuts

Complementary therapy, placebo, and odd techniques make up most of ***Bandolier***'s questions. So good information that doing nothing doesn't make things happen, that acupuncture doesn't work (this time with cocaine addiction), and a real puzzler over pulsed ultrasound and fracture healing. Statistical significance on tiny numbers is not always a good reason for doing something. The frustration is that no-one is helping us deal with the common situation of too little information.

TNF ANTIBODIES AND RHEUMATOID ARTHRITIS

A real difficulty with original treatments is that evidence emerges slowly. This is particularly the case for chronic diseases where treatments may have to be assessed over long periods. This slow emergence means that there can be some initial difficulty in assessing whether the treatment works, how well it works, in whom it works, and what the economics are if it does work.

There comes a point, though, when there is sufficient evidence on which to pass a judgement, which is one reason the UK government set up its National Institute for Clinical Excellence (NICE). As part of its work, NICE now publishes the evidence on which its judgements are based on its Internet site. The way in which it assessed the evidence on anti-TNF treatments for rheumatoid arthritis [1] demonstrates a growing quality and maturity in its processes.

NICE evidence

The 138-page report from Birmingham is a good background read on rheumatoid arthritis and its treatment. It examines the background, diagnosis, pathology, current service provision, the interventions, its search strategy, the results in the form of a meta-analysis of major outcomes, detailed summaries of the trials involved including adverse events, and the health economic arguments, as well as the implications for the NHS.

Technology

Infliximab and etenercept are partially humanised monoclonal antibodies aiming to reduce the actions of circulating tumour necrosis factor (TNF). Both exert their effects by removing TNF from the circulation, and consequently interrupt the inflammatory process.

In this issue

Efficacy

With rheumatoid arthritis there are many possible outcomes, starting with the number of painful or swollen joints, through to health assessment questionnaires and quality of life. The most commonly used (though perhaps difficult to explain and use) measure is the American College of Rheumatology (ACR) response criteria, a composite measure of seven indices:

- Tender joint count
- Swollen joint count
- Global disease activity assessed by observer
- Global disease activity assessed by patient
- Patient assessment of pain
- Physical disability score (like health assessment questionnaire)
- Acute phase response (CRP measurement, or ESR)

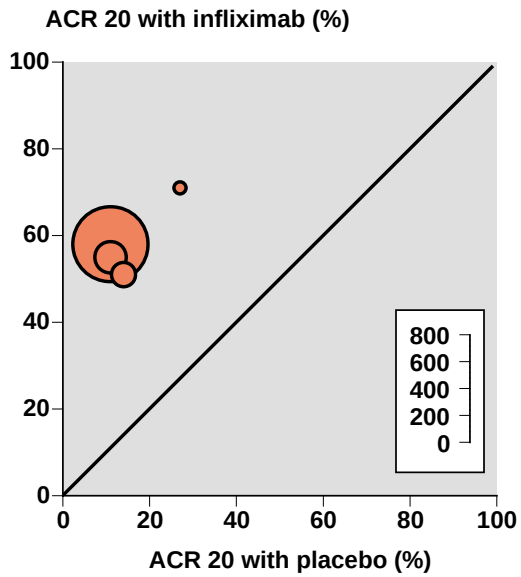
The ACR 20 response is defined as a 20% improvement in the first two of these, plus a 20% improvement in any three of the remaining five items. This is not an easy outcome to reach, though also being used now are ACR 50 and ACR 70, which are similar to the ACR 20 but at 50% and 70% improvements. These are very high hurdles of treatment efficacy and represent very significant clinical improvement.

Results

To at least give a flavour of the results obtained with these two monoclonal antibodies, the results for ACR 20, 50 and 70 are looked at here. There are issues about dose, and about duration. The NICE evidence combines different doses and duration up to one year, but shows that, in the main, this is a reasonable thing to do, at least for now.

Results were consistent between trials comparing the new treatments with placebo. Those for the ACR 20 for infliximab are shown in Figure 1, and for etenercept in Figure 2. The results for ACR 20, 50 and 70 outcomes are in Table 1. NNTs were about 2-3 for ACR 20, about 4 for ACR 50, and about 8 for ACR 70.

Figure 1: ACR 20 with infliximab and placebo



The analysis also shows that what happens with placebo depends on the difficulty of the outcome. With placebo about 14% of rheumatoid arthritis patients achieved ACR 20. As the hurdle was raised to ACR 50 and 70, the percentage of patients achieving the outcome fell to about 6% and 1% respectively (Table 1).

Good information about adverse events is also given in the report, though not recounted here.

Comment

Given the evidence outlined in the report, it is clear why the British Society for Rheumatology recommended that infliximab and etenercept should be used if the following criteria were met:

- Patients satisfy ACR classification for RA
- Patients have highly active RA
- Patients should have failed treatment on methotrexate and at least one other disease modifying agent
- Treated patients should be entered on a central register, with drugs, dose, outcomes and toxicity reported on a quarterly basis.

Early referral

An essential accompaniment to the NICE evidence is an evidence-based early referral recommendation for newly diagnosed rheumatoid arthritis [2]. Based on a wide literature search by an international group of influential rheumatologists, this review provides clear support for the observation that permanent structural damage occurs early in the course of active rheumatoid arthritis and that early disease modifying drug intervention slows the progression of structural joint damage and improves long term outcomes, as well as overall quality of life.

It provides an early referral algorithm for newly diagnosed rheumatoid arthritis. Early referral is advised in the event of clinical suspicion of RA, supported by the presence of any of the following:

Figure 2: ACR 20 with etenercept and placebo

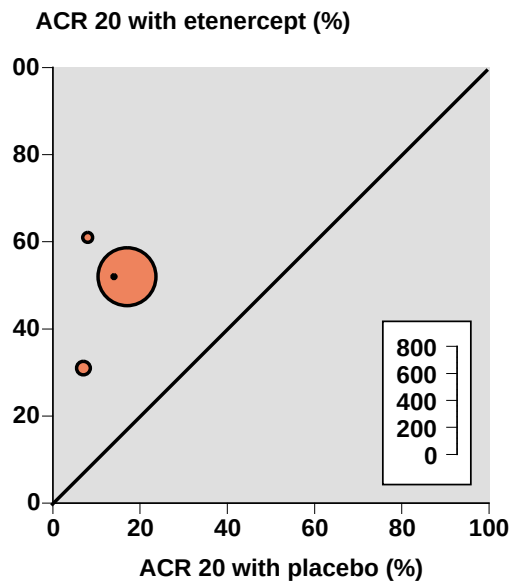


Table 1: NNTs for infliximab and etenercept versus placebo at longest duration, and pooling all doses

Treatment	Outcome	Trials	Outcome [number/total (%)] with Treatment	Outcome [number/total (%)] with Placebo	Relative benefit (95%CI)	NNT (95%CI)
Infliximab	ACR 20	4	458/803 (57)	35/259 (14)	4.3 (3.1 to 5.9)	2.3 (2.0 to 2.6)
	ACR 50	4	232/803 (29)	14/259 (5)	5.5 (3.3 to 9.2)	4.3 (3.6 to 5.2)
	ACR 70	3	68/667 (10)	2/215 (1)	12 (3 to 47)	11 (8 to 15)
Etenercept	ACR 20	4	245/497 (49)	19/133 (14)	3.6 (2.3 to 5.4)	2.9 (2.4 to 3.6)
	ACR 50	4	161/497 (32)	10/133 (8)	4.5 (2.4 to 8.3)	4.0 (3.2 to 5.3)
	ACR 70	1	60/340 (18)	2/88 (2)	8 (2 to 31)	6.5 (4.9 to 9.7)

1 At least three swollen joints
2 A positive "squeeze" test (composite compression test) that clinically evaluates a group of small adjacent joints such as metacarpophalangeal and metatarsophalangeal
3 Morning stiffness of at least 30 minutes

Important points were:

- Patients with RA have been shown to have an improved long term outcome, when treated by a rheumatologist
- There is evidence that delay of more than 12 weeks in treatment results in a missed opportunity to improve long term outcome
- RF positivity, a raised acute phase response, and erosions on X-ray are associated with poor outcome, but absence at presentation should not preclude diagnosis or referral
- NSAIDs can mask signs and symptoms at presentation
- Corticosteroids should not be prescribed without an accurate diagnosis.

The algorithm combines all these points with a diagram of the squeeze test, and it would make a useful guide for teaching, or a surgery wall, or computer aide.

Overall comment

There is much new in the field of rheumatoid arthritis, and probably much more to come, backed by good new evidence. Given that RA is the most common form of inflammatory arthritis affecting 0.5% to 1% of the population, and with a large economic impact because about 90% of patients have some form of disability with two decades, this is of increasing importance. Some or all of this could be incorporated into the delivery of a quality service.

A short review of new treatments can also be found in the STEER series from the Wessex Institute [3].

References:

1 P Jobanputra et al. The clinical effectiveness and cost-effectiveness of new drug treatments for rheumatoid arthritis: etenercept and infliximab. NICE 2002 (http://www.nice.org.uk/Docref.asp?d=29675)
2 P Emery et al. Early referral recommendation for newly diagnosed rheumatoid arthritis: evidence based development of a clinical guide. Annals of Rheumatic Disease 2002 61: 290-297.
3 C de Vries. Effects of TNF-alpha antagonists in people with rheumatoid arthritis. (http://www.signpoststeer.org)

THAT OLD PLACEBO FEELING

Bandolier continuously gets into trouble over placebo because people have their own very fixed views about what it is. The extremes of definition are shown in Figure 1.

One definition is that use of a placebo describes what happens when you do nothing, so that in the context of a clinical trial, for instance, a placebo group could describe the natural history of a disorder without the intervention under test. It is the effect we see when we do nothing.

This rather clumsy phrase is abbreviated to the *placebo effect*. The trouble is that placebo effect can be taken to mean that any effect seen is ***caused by*** placebo. The definition has now changed to mean that doing nothing causes things to happen. The idea that placebo ***does*** make things happen is reinforced by much interesting academic work, almost always in contrived situations and with small numbers of people. Extrapolating that to clinical practice is a big step, but one often assumed.

Is it possible to say that one definition is right and the other wrong? A systematic review and analysis comparing clinical trials in which patients were randomly assigned to placebo or no treatment suggests that the evidence for placebo having an effect is at best weak [1].

Placebo:
what happens when you do nothing

Placebo effect:
doing nothing makes things happen

Review

The review sought studies comparing placebo with no treatment that were properly randomised and blinded. Exclusions were healthy volunteers, studies with high dropout rates, or if the alleged placebo had some clinical effect.

Results

The analysis included 114 trials in 40 clinical conditions.

For dichotomous outcomes in 32 trials with 3,800 patients (average 119 per comparison) there was no difference between placebo and no treatment overall, or with subjective or objective outcomes.

For continuous outcomes in 82 trials with 4,800 patients (average 59 patients per comparison) there was a small overall effect, all for subjective outcomes.

The only condition in which there was an effect was for pain for continuous outcomes in 27 trials with 1,600 patients (average 59 trials per comparison).

Comment

The interesting thing is that any possible effect of placebo was found with continuous outcomes for pain in small trials. This is likely to be an artefact. Firstly we know [2] that continuous pain outcomes are often reported as a mean of data with highly skewed distribution. Analysis of means is therefore meaningless [2]. Smaller trials in pain have also been shown to have more effect than larger ones, probably because they are of poorer quality [3].

Not much evidence there, then, that doing nothing makes things happen.

References:
1 A Hröbjartsson & PC Gøtzsche. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. New England Journal of Medicine 2001 344: 1594-1602.
2 HJ McQuay et al. Variation in the placebo effect in randomised controlled trials of analgesics: all is as blind as it seems. Pain 1996 64:331-335.
3 RA Moore et al. Quantitative systematic review of topically-applied non-steroidal anti-inflammatory drugs. British Medical Journal 1998 316: 333-8.

ACUPUNCTURE FOR COCAINE ADDICTION

A big problem with complementary therapy is that so many of the trials are of such low quality that bias of some degree is likely. When a properly done trial of high quality comes along, it therefore deserves our attention. Such a study demonstrates that acupuncture is ineffective for cocaine addition [1].

Trial

This trial compared acupuncture of the ear using criteria of the US National Acupuncture Detoxification Association with needles inserted into non-acupuncture points in the ear and a non-acupuncture relaxation regimen for eight weeks. Six hundred and twenty cocaine dependent adults were randomised. They had to have a recent cocaine-positive urine screen and self-reported cocaine use, but not be dependent on any other drug except nicotine or opiates, or currently receiving other treatment, be actively suicidal or psychotic, or received acupuncture in the previous 30 days.

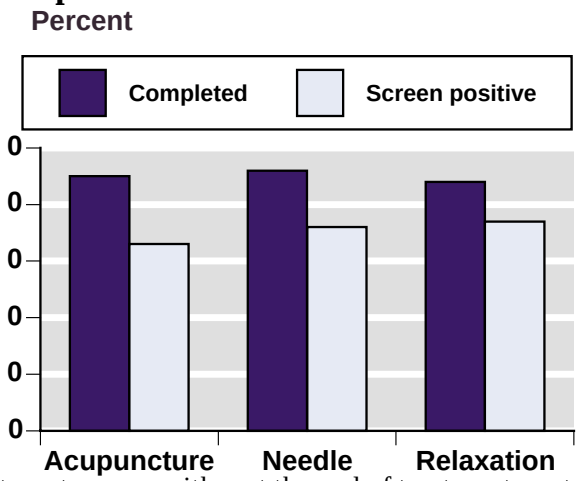
Patients were told their treatment assignment, because blinding was not possible. Treatments were described in a standardised way, with all the treatment options described as ways to reduce stress and with potential benefits to reduce craving and cocaine use. Weekly (standardised) individual counselling sessions focusing on changing addictive behaviour were offered as adjunctive treatment.

Outcomes were positive urine screens, self-reported cocaine use, and a range of other outcomes.

Results

The three groups were identical in terms of age, sex and ethnic background. Over the eight weeks of the study retention was under 50% and equal in the three groups (Figure 1). The proportion of cocaine-positive urine tests (averaging about three a week) was the same in the three groups (Figure 1). While there was a reduction in cocaine use over the eight weeks, there was no difference between the three

Figure 1: Percentage of randomised patients who completed eight weeks, and percentage of urines positive for cocaine



treatment groups, either at the end of treatment or at three or six months of follow up (Figure 2).

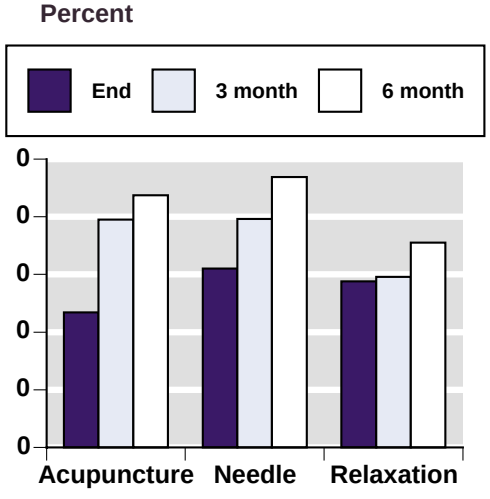
Comment

This is a good study that looks at the result both by intention to treat and by completers. It failed to find any difference between acupuncture recommended by the key US association, sham acupuncture, and the non-acupuncture control of relaxation. It dealt with the fact that blinding was impossible, it was randomised, and it was large. The lack of blinding could cause bias, but should have been in the direction of acupuncture doing better if patients or therapists believed that it would. There was no difference.

Acupuncture doesn't work on the basis of these results. What the paper does tell us is how to maximise methodological rigour in a difficult clinical situation, and helps set standards against which we can compare small, low quality studies that more often say that acupuncture does work. It helps explain why small, low quality studies in acupuncture, or anything else, should be ignored.

References:
1 A Margolin et al. Acupuncture for the treatment of cocaine addiction. JAMA 287: 55-63.

Figure 2: Percentage of patients completing the study who were abstinent at the end, and after three and six months



PULSED ULTRASOUND FOR FRACTURE HEALING

Some fractures have delayed healing or nonunion. Ultrasound has been suggested as a possible therapeutic agent to improve healing. Does it work? A new systematic review and meta-analysis shows how little evidence we have [1].

Search

This was impressive, using many databases and attempts to find unpublished material, plus hand searching key journals. Methods used were great, and clearly defined inclusion criteria were random allocation to pulsed ultrasound or control, skeletally mature patients, double blinding, with time to radiological healing as the main outcome.

Results

There were three trials involving 158 fractures (tibial shaft, distal radius and scaphoid, respectively). The mean times to healing are shown in Figure 1. Mean healing time was reduced by about 60, 40 and 20 days respectively.

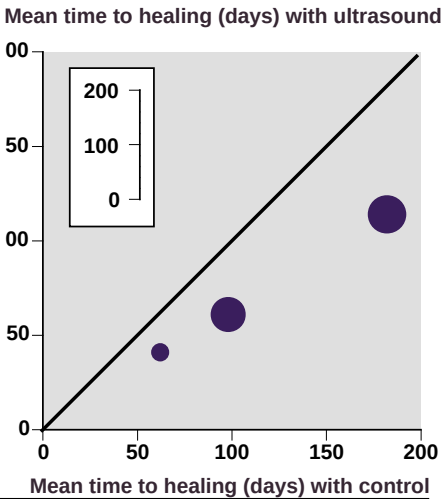
Comment

The authors are justifiably cautious in their conclusions. They say that pulsed ultrasound ***may*** reduce time to fracture healing for fractures treated nonoperatively. They imply that we should beware because of the small amount of data. And the trials are good, with high quality scores, so this isn't like some reviews with small amounts of poor information that ladle on heaps of weasel word sauce.

What we frequently lack in systematic reviews is some idea of ***how much*** information we need to be sure of the result. If there is little information, as here, then unpublished negative trials could be very important. Someone will be selling this to you soon on the basis that it is "evidence-based"!

Reference:
1 JW Busse et al. The effect of low-intensity pulsed ultrasound therapy on time to fracture healing: a meta-analysis. Canadian Medical Association Journal 2002 166: 437-441.

Figure 1: Healing time with ultrasound and placebo



Depression is awfully common. In primary care, estimates of the prevalence of major depression range from 5% to 9%, and that's just the patients. We all get fed up, we all get tired, we all get to feel a bit worthless from time to time. We may not be well, or be recovering from the last in a succession of viruses. Sometimes we pull ourselves together, and sometimes we can't. The problem for primary care professionals is telling major depression from being fed up, and then whether that depression signifies some unrecognised problem.

A terrific new systematic review [1] pulls the subject together beautifully, and gives us a clear insight into the latest diagnostic criteria for depression plus good information on the performance of various simple screening and diagnostic systems for primary care.

Diagnostic features

Based on the criteria from Diagnostic and Statistical Manual of Mental Disorders (DSM-IV), eight symptoms and criteria are defined, together with some suggested questions for patients (Table 1). Depending on the answers to the questions and the duration of the symptoms depression can be major or minor, or can just indicate dysthymia (melancholy, Table 2). Can diagnostic questionnaires be of any help?

Table 2: Diagnostic categories for depression and dysthymia (melancholy), and criteria for diagnosis

Diagnostic category	DSM-IV criteria	Duration
Major depression	≥5 depressive symptoms, including depressed mood or anhedonia, causing significant impairment in social, occupational, or other important areas of functioning	≥2 weeks
Minor depression	2 to 4 depressive symptoms, including depressed mood or anhedonia, causing significant impairment in social, occupational, or other important areas of functioning	≥2 weeks
Dysthymia	3 or 4 dysthymic symptoms, including depressed mood, causing significant impairment in social, occupational, or other important areas of functioning	≥2 years

Table 1: DSM-IV diagnostic criteria and suggested questions

Symptom	DSM-IV diagnostic criteria	Suggested questions
Depressed mood	Depressed mood most of the day, nearly every day	How has your mood been lately? How often does this happen? How long does it last?
Anhedonia	Markedly diminished interest or pleasure in almost all activities most of the day, nearly every day	Have you lost interest in your usual activities? Do you get less pleasure in things you used to enjoy?
Sleep disturbance	Insomnia or hypersomnia nearly every day	How have you been sleeping? How does that compare with your normal sleep?
Appetite or weight change	Substantial change in appetite nearly every day or unintentional weight loss or gain (≥5% of body weight in a month)	Has there been any change in your appetite or weight?
Decreased energy	Fatigue or loss of energy nearly every day	Have you noticed a decrease in your energy level?
Increased or decreased psychomotor activity	Psychomotor agitation or retardation nearly every day	Have you been feeling fidgety or had problems sitting still? Have you slowed down, like you were moving in slow motion or stuck in mud?
Decreased concentration	Diminished ability to think or concentrate, or indecisiveness, nearly every day	Have you been having trouble concentrating? Is it harder to make decisions than before?
Guilt or feelings of worthlessness	Feelings of worthlessness or excessive guilt nearly every day	Are you feeling guilty or blaming yourself for things? How would you describe yourself to someone who had never met you before?
Suicidal ideation	Recurrent thoughts of death or suicide	Have you felt that life is not worth living or that you'd be better off dead? Sometimes when a person feels down or depressed they might think about dying. Have you been having any thoughts like that?

Table 3: Diagnostic utility of case-finding instruments for depression in primary care

Instrument	Number of Studies	Number of Patients	Likelihood ratio Positive	Likelihood ratio Negative
Centre for Epidemiological Studies depressive screen	10	3038	3.3	0.2
Symptom driven diagnostic system	4	1682	3.5	0.2
Beck depression inventory	4	952	4.2	0.2
Zung self assessment depression scale	4	667	3.3	0.4
Primary care evaluation of mental disorders	2	967	2.7	0.1
Hopkins symptom check list	2	946	3.2	0.2
Geriatric depression scale	2	165	3.3	0.2
Patient health questionnaire	1	585	12	0.3
Single question	1	291	2.3	0.2

Search

MEDLINE and the Cochrane register of depression trials were searched for English-language studies evaluating the performance of case-finding instruments in primary care and the reliability of the clinical interview. Case-finding instruments had to have easy to average literacy requirements, be scored without a calculator, have a depression-specific component and be evaluated in at least one study with at least 100 subjects. Reliability studies had to have diagnoses made by two or more clinicians reviewing audio- or videotape interviews.

Results

Eleven questionnaires with between one and 30 items were found in 28 studies involving over 25,000 patients for case-finding. Major selection bias occurred in nine trials, which were then excluded. Three instruments, Beck depression inventory, Centre for Epidemiological Studies depressive screen and the Zung self-assessment depressive scale were developed specifically to identify depression. None of the instruments took more than five minutes to administer, and most could be done in two or three minutes.

The results for finding major depression in a primary care setting are shown in Table 3, with the mean likelihood ratios for a positive and a negative result. The likelihood ratio for a positive result was about 3 and for a negative result was about 0.2. In a clinic with an 8% prevalence of major depression of dysthymia, a clinician seeing 100 patients a week can expect that 30 will screen positive for depression, of whom seven would meet the criteria for major depression after a more careful interview. In the 70 patients who screen negative, one would actually be clinically depressed.

In seven studies using a semistructured interview, agreements between examiners was good, with kappa values generally above 0.7.

Comment

This is a splendid review, with much more useful information than can be conveyed by précis. For any primary care organisation or professions wanting to improve the determination of depression in a primary care setting reading this is an essential start. The ***Bandolier*** Internet version has a set of links to the various instruments used in the study. Not all of them have a website, but of those that do, the commonest are shown in the box below.

References:
1 JJ Williams et al. Is this patient clinically depressed? JAMA 2002 287: 1160-1170

Websites for common instruments

Centre for Epidemiological Studies at
chipts.ucla.edu/Assessment_Instument/asmt_dp2.html
Beck Depression Inventory at
www.uea.ac.uk/~wp316/depression.pdf
Zung at:
www.wellbutrin-sr.com/hcp/depression/zung.html
Primary care evaluation
(PDF by email to Robert Spitzer on RLS8@colombia.edu)
Geriatric Depression Scale at
www.stanford.edu/~yesavage/GDS.english.long.html
Patient Health Questionnaire at
www.cmecenter.com/primemtoday/

Note that some deep links may not work, and you may have to use just the first part of the Internet address

EDITORS

Andrew Moore Henry McQuay
Pain Relief Unit
The Churchill, Oxford OX3 7LJ
Editorial office: 01865 226132
Editorial fax: 01865 226978
Email: andrew.moore@pru.ox.ac.uk
Internet: www.ebandolier.com
ISSN 1353-9906

BANDOLIER SUBSCRIPTION FORM

Important
All existing non-subscription distribution will end in July 2002 and from August 2002 Bandolier will be available only by subscription direct from us. If you wish to continue to receive Bandolier after July you have to complete this form and return it to us in good time.

Title you want us to use	
Name	
Address	
Address	
Address	
Postcode and country	
Email	
Phone or fax	**Tel** **Fax**
Address to send Bandolier if different from above	

I would like to subscribefor 12 issues **UK & N Ireland £36** ☐ **Rest of the world £72** ☐

Payment options (Pounds Sterling only) including postage

Cheque (UK only) ☐ **International Money Order (RoW & Ireland)** ☐

Please make cheques payable to Bandolier

Visa ☐ **Mastercard** ☐

Name on card

Card number

Expiry date

Signature **Date**

Please send the completed form by fax to +44 (0)1865 226978, or with payment, to:

Maura Moore
Bandolier Subscriptions
Pain Research
The Churchill
Oxford OX3 7LJ, UK

Bandolier would be delighted to quote for bulk orders from organisations, especially healthcare institutions or charities, or for bulk reprints, all at preferential rates: Email: maura.moore@pru.ox.ac.uk